A Molecular Approach to Carcinogenesis: Dissecting the Genome of a Small Tumor Virus*,†

DANIEL NATHANS

Department of Microbiology, Johns Hopkins University School of Medicine, Baltimore, Maryland

Received for publication July 15, 1976

Transformation of a normal cell to a cancer cell involves a stable, heritable change in the growth pattern of that cell and its descendants: the cancer cell proliferates under conditions where the normal parental cell does not. In simplistic molecular terms we would hypothesize that a primary change has occurred in one or more cellular genes, and/or in the expression of certain genes, which leads perhaps through many intermediate steps to loss of growth control. For example, most chemical carcinogens are known to be potent mutagens, and their carcinogenicity is thought to be due to mutation. However, neither sites of mutation nor the links between mutation and growth alteration are known. Other tumorigenic agents are certain kinds of viruses. In this case the primary event is the introduction of new genetic information into the cell, namely, the cluster of genes which make up the nucleic acid genome of the virus. Depending on the genetic content of the virus and on the circumstances of infection, expression of viral genes can lead to multiplication of virus and cell in which the cell acquires new heritable properties including, in some cases, altered growth in culture and the ability to form a tumor in an animal. In these "transformed cells' viral genes persist, often as part of cellular chromosomes, and it is likely that specific viral gene products are continuously needed to maintain the tumorigenic properties of the cell. Since viruses are rather simple and can be readily manipulated genetically and biochemically, a great deal of attention has been focused on the molecular biology of tumor viruses in the expectation of better understanding tumorigenesis in general and also associated fundamental mechanisms of gene action and growth regulation in animal cells.

Tumor viruses are not basically different from other viruses. While only certain types of RNA viruses are tumorigenic, almost every type of DNA containing animal virus is potentially tumorigenic. Among the simplest of these are the small papovaviruses, including Simian Virus 40 (SV40), the murine polyoma virus, and the human papovaviruses BK and JC. Owing to their small size, ease of handling, and possible role in human disease, these viruses are being intensively investigated in a number of laboratories. One line of investigation, including our own studies, has centered on the molecular genetics of SV40, and in particular on mapping the SV40 chromosome by enzymatic dissection and restructuring of the viral DNA. The result has been the construction of a physical map of the SV40 genome and localization of template functions, structural genes, and genetic determinants of biological functions within the viral chromosome. We now have a rather detailed picture of the organization of the SV40 genome, including that part of the molecule that is responsible for tumorigenicity. (For a review, see Ref. 1).

The Structure of SV40 and Its DNA

SV40 occurs as an inapparent natural infection of certain species of Asiatic macaques (2). Similar papovaviruses have since been detected in man. From an experimental viewpoint the attractive features of SV40 are its especially simple and rather stable structure, and its small genome. As shown in Fig 1, the virion or virus particle of SV40 is a small icosahedron. Each virus particle contains a molecule of covalently closed circular duplex DNA of about 3.2×10^6 daltons. As shown in Fig 1, DNA molecules of this type (called Form I DNA) have supercoils and are therefore more compact than duplex circles con-

* The research carried out in the author's laboratory was supported by grants from the National Cancer Institute (CA-16519), the American Cancer Society (VC-132A), and the Whitehall Foundation.

† Presented as the Dean's Lecture, The Johns Hopkins University School of Medicine, February 2, 1976.

Reprints requests to Dr. Daniel Nathans, 725 North Wolfe Street, Baltimore, Maryland 21205

THE JOHNS HOPKINS MEDICAL JOURNAL **139** 99-120 (1976)

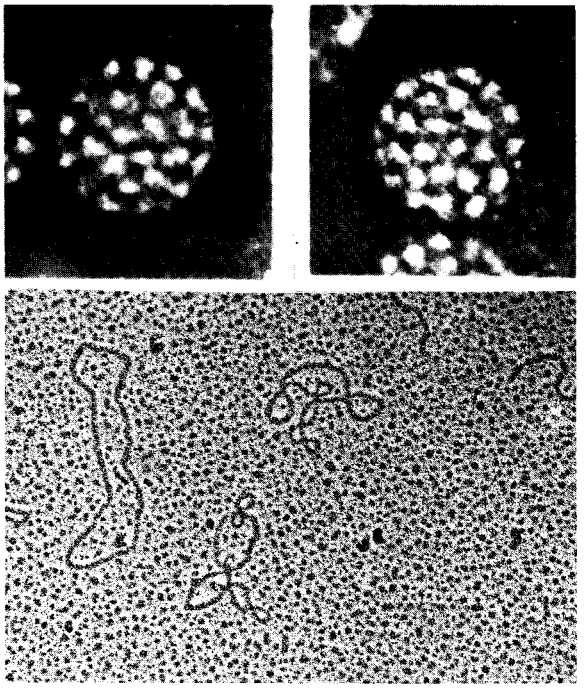

Fig. 1. Electron micrographs of SV40 virions (top) and SV40 DNA (bottom). The micrographs above were reproduced with permission from Anderer et al. (3), and that below was provided by N. Muzyczka. (The linear DNA molecules resulted from cleavage of form I DNA by restriction endonuclease.)

taining one or more single-strand breaks (called Form II DNA), also illustrated in Fig 1. The relationship between Form I and Form II DNA is shown better in Fig 2, which illustrates the covalently closed supercoiled duplexes and the result of creating a break in one of the two DNA strands or across both strands. From electron microscopic length measurements of the DNA (4) one can estimate that the number of nucleotide pairs in an SV40 DNA molecule is approximately 5000, which is sufficient information to code for about 1700 to 1800 amino acids.

SV40-cell Interactions

In spite of its paucity of genetic information, SV40 can replicate in the absence of a helper virus in monkey cells, and can also transform to tumorigenicity other types of cells in which it does not replicate, for example, mouse cells. Thus, SV40 provokes two distinctly different types of cell reactions: productive infection and cell death on the one hand, and transformation to tumorigenicity on the other. Some of the molecular events involved in these two types of virus-cell interactions are outlined diagrammatically in Fig 3. As seen on the left side of Fig 3, productive infection begins with virus absorption to the cell surface followed by appearance of viral DNA molecules in the cell nucleus. Certain viral genes called "early" genes are then transcribed in the cell nucleus into early messenger RNA which is translated in the cytoplasm into early protein. Appearing in the nucleus at this time is a new, SV40-specific antigen (T or tumor antigen) which is now thought to be the early protein. Early protein acts in the nucleus in the initiation of viral DNA replication. Following viral DNA replication other genes of the viral chromosome are transcribed into messenger RNA, these are the "late" genes giving rise to late messenger RNA. The late messenger RNA is translated into structural proteins of the virus which then form the outer shell or coat of new virus particles. Eventually a large number of such particles are produced in the cell, leading to cell lysis and release of virus particles into the medium.

In contrast to productive infection, transformation by SV40 involves *partial* expression of viral genes. As shown on the right side of Fig 3, when the viral DNA reaches the nucleus of cells destined to be transformed, genes are expressed in the form of messenger RNA and early viral protein (T antigen), but viral RNA replication does not ensue. Instead, a molecule of viral DNA becomes covalently linked or integrated into the cellular DNA as diagrammed in Fig. 3, following which, viral genes continue to be expressed in the form of messenger RNA and viral

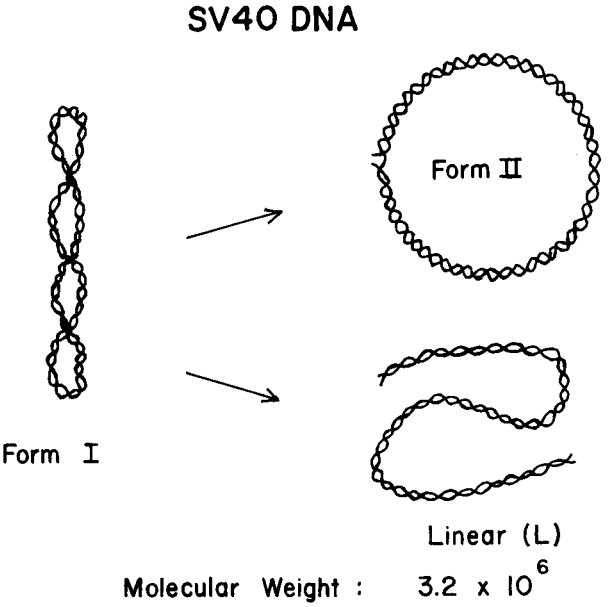

Fig. 2. A diagram of SV40 DNA form I (covalently closed circular duplex), form II (nicked or open circular duplex), and linear DNA (L) after double-strand scission.

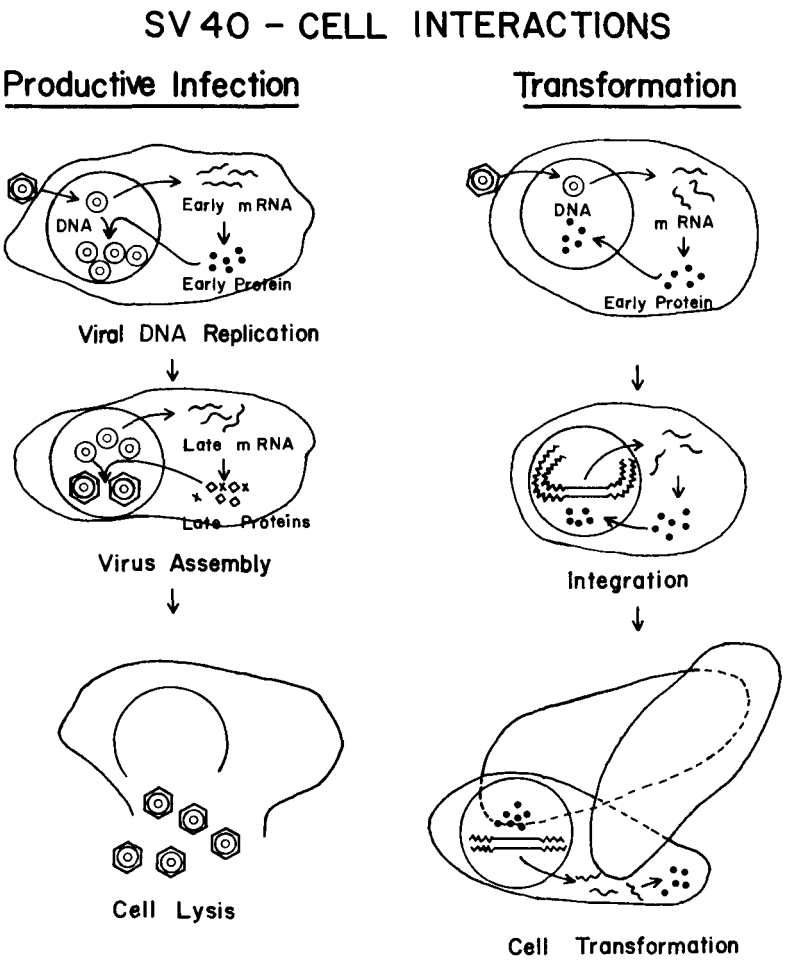

Fig. 3. Diagram of SV40-cell interactions. Among the virus-specific proteins found in infected permissive cells and in transformed cells is the SV40 T-antigen. See text for discussion.

protein. Integration of the viral genome thus creates a stable condition in which viral genes persist in the transformed cell and are inherited by each daughter cell. In such cells viral messenger RNA continues to be made, as does viral protein. As described later, the continued expression of these genes is required to maintain the transformed state.

The objective of the studies to be described has been to determine how the SV40 genetic information expressed in productively infected or transformed cells (as diagrammed in Fig 3) is organized: the number and location in the viral chromosome of early and late genes and those genes transcribed in transformed cells; the identity of specific gene products; the direction of gene transcription; whether there are specific signals for viral DNA replication and, if so, where these are located; and finally, which genes are essential for transformation.

CLEAVAGE MAP OF THE SV40 GENOME

Our first aim in delineating the genetic organization of SV40 was to construct a physical map of the viral DNA molecule based on specific cleavage of the DNA by site-specific bacterial restriction endonucleases. Restriction endonucleases are a class of enzymes found in many bacterial species, which cleave foreign DNA but not the DNA of the cell that contains the enzyme (for recent review, see Ref. 8). Some enzymes of this type cleave DNA at specific nucleotide sequences, as first shown by my colleagues, Thomas J. Kelly, Jr. and Hamilton O. Smith (5) with an enzyme isolated from *Hemophilus influenzae* (6). Thus, these enzymes are analogous to specific proteolytic enzymes such as trypsin and chymotrypsin which cleave proteins at specific amino acid residues.

We began our analysis of SV40 when Stuart Adler, at that time a Hopkins medical student, surveyed the known restriction enzymes for their ability to cleave SV40 DNA (7). Most useful, it turned out, was Smith's enzyme from *H. influenzae*. Since those initial studies, many other site specific restriction enzymes have become available which also cleave SV40 DNA (for review, see Ref. 8). The nucleotide sequence recognized by several enzymes active on SV40 DNA are shown in Fig 4. As seen in the figure, each enzyme recognizes a specific base sequence in duplex DNA and cleaves both strands of the duplex as shown by the arrows. Each recognition site consists of a symmetrical sequence of nucleotides in which the base sequence in one strand is the same as that in the opposite strand. Some enzymes cleave duplex DNA evenly, i.e., the resulting fragments are even-ended, whereas other enzymes make staggered breaks in DNA resulting in fragments with short, single-strand tails. Such single-strand tails or "sticky ends" allow ready joining of one fragment to another by base pairing, or cyclization of a given fragment into a circle (Fig 5). These properties are the basis for certain recombinant DNA experiments and, as will be described later, allow the construction of excisional deletion mutants of SV40.

The general strategy of the approach taken by my coworkers, K. J. Danna and G. H. Sack, Jr. for constructing a cleavage map of the SV40 DNA molecule was to cleave DNA with one or more restriction enzymes, separate the resulting fragments by gel electrophoresis and determine the size of each fragment and its position relative to the others in the original DNA molecule (9, 10). As illustrated in Fig 6, SV40 DNA can be cut into a number of smaller linear fragments by a particular restriction enzyme, thus giving rise to segments of DNA from specific sites in the genome. On the other hand, with other enzymes the circular molecule is cut once at a specific site to yield a full length linear molecule of SV40 DNA (11, 12). Figure 7 illustrates the electrophoretic separation of SV40 DNA fragments which result from cleavage by a series of restriction endonucleases. Since the mobility of DNA fragments is proportional to length, one can with a suitable series of standards estimate the length of a given DNA fragment by its electrophoretic mobility. Alternatively, one can determine length by the amount of DNA contained in

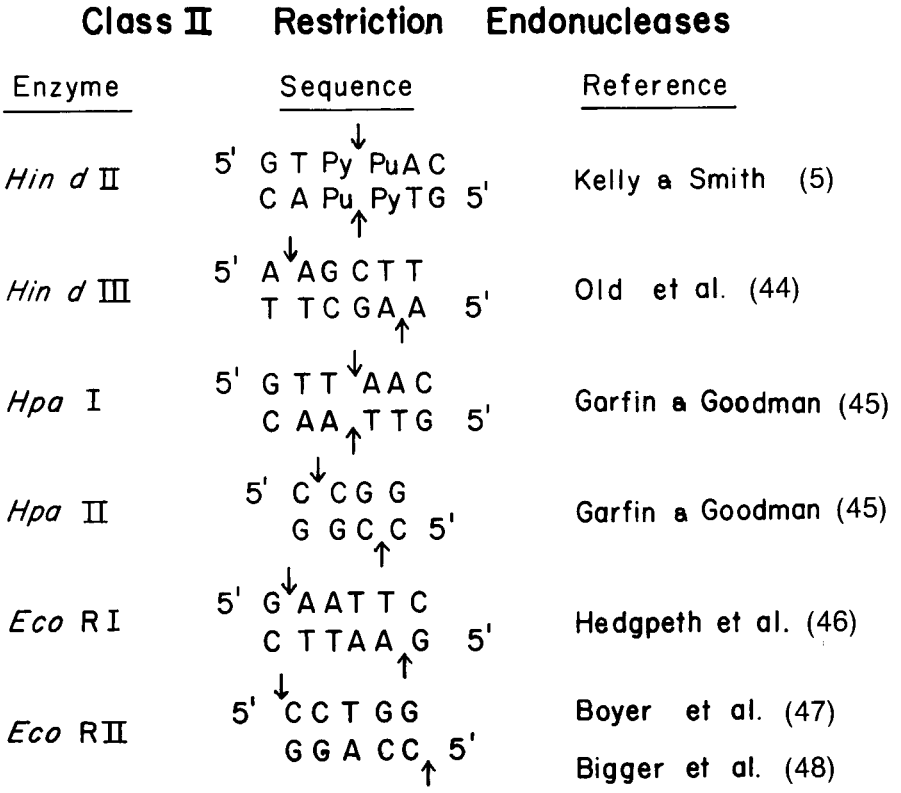

Fig. 4. Examples of bacterial restriction endonucleases that cleave SV40 DNA and the nucleotide sequences recognized. Arrows indicate sites of cleavage. Hin, *H. influenzae*, Hpa, *H. parainfluenzae*, Eco, *E. coli* (8).

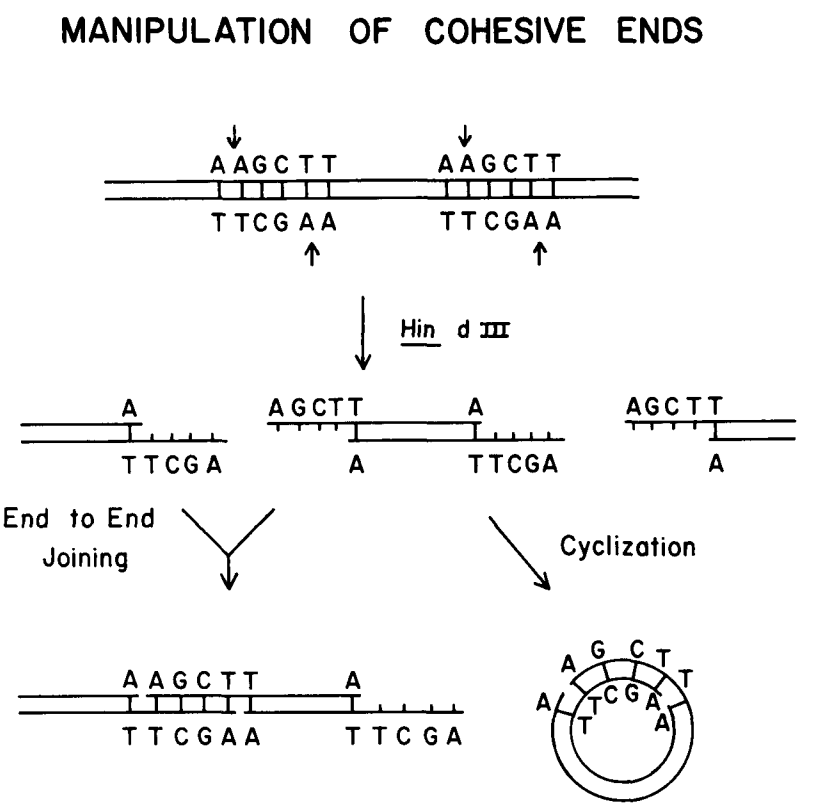

Fig. 5. The use of cohesive ends in DNA (produced by the *Hin*dIII restriction endonucleases in this case) to join or cyclize DNA fragments.

each fragment relative to the starting DNA molecule or by electron microscopic length measurements vs the length of the original SV40 DNA.

To determine the order of a given set of fragments in the original DNA molecule several methods have been used, notably the analysis of partial digest products, as illustrated in Table I. In this instance, SV40 DNA was digested with a small amount of *Hin* enzyme and the partial digest products isolated after electrophoresis in acrylamide gel; each partial digest product was then digested to completion with the same enzyme and the digest products analyzed by gel electrophoresis. As shown in the table, each set of final digest products can be arranged in a consistent overlapping order to result in the order of each of the fragments in the original circular SV40 DNA molecule. The fragments so arranged then constitute the cleavage map, as shown in Fig 8. Similar analyses of enzyme digest products have been carried out in a number of laboratories with many different restriction enzymes resulting in the detailed cleavage map of the SV40 DNA molecule shown in Fig 9.

Having such a map we could now proceed to localize template functions, structural genes, and

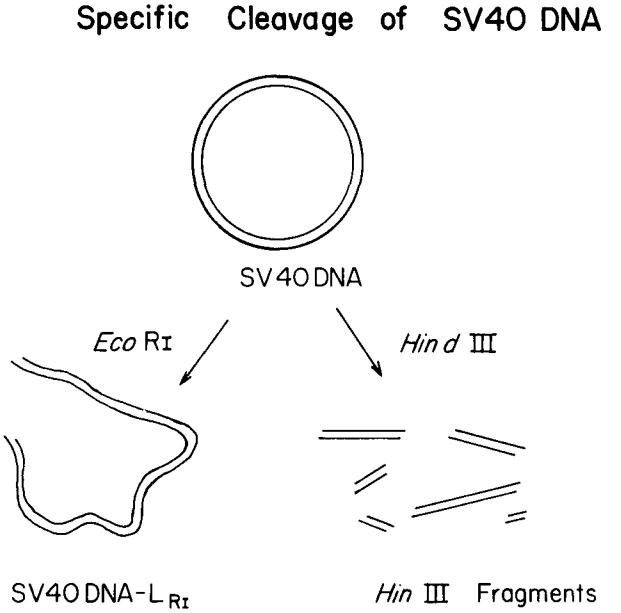

Fig. 6. Diagrammatic representation of action of restriction endonucleases on SV40 DNA. In one case (left) the DNA is cut once at a unique site and in other (right) it is cut at several specific sites.

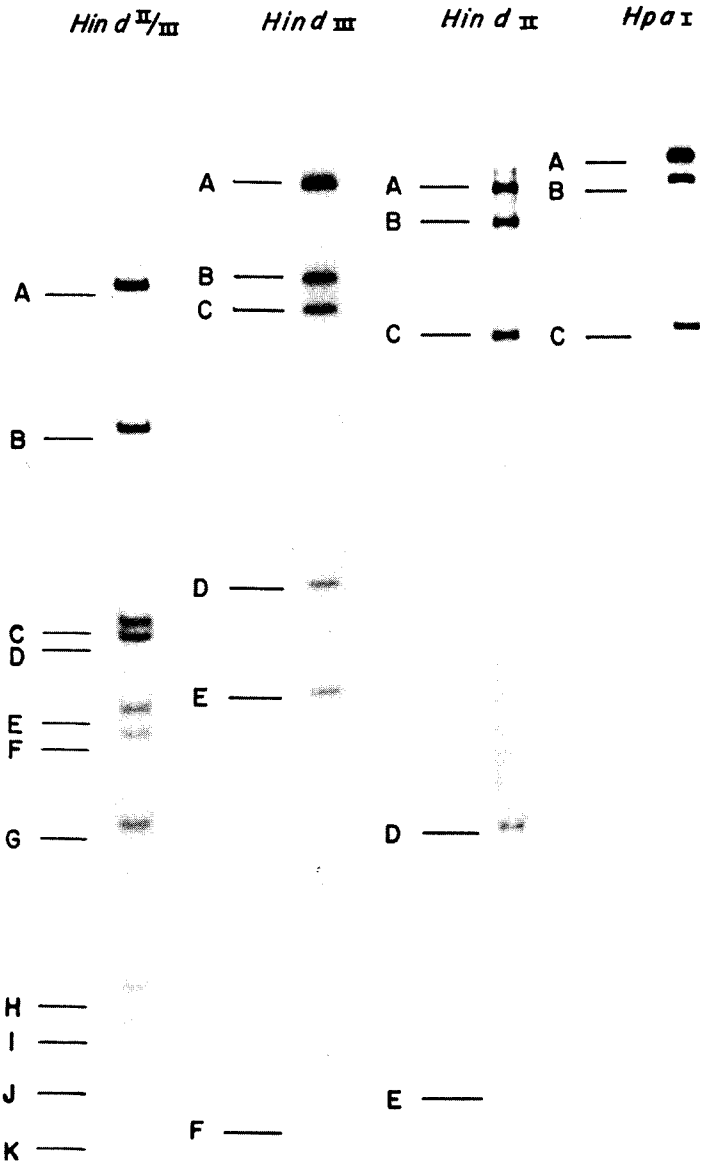

Fig. 7. Fragments of SV40 DNA produced by various restriction endonucleases and separated by electrophoresis in polyacrylamide gels: autoradiograms of ^{32}P-labeled fragments. The origin is at the top. Reproduced, with permission, from *The Harvey Lectures,* Series 69, p. 117, 1974–1975.

biological activities of the SV40 genome in the cleavage map.

MAPPING TEMPLATE FUNCTIONS

The first sites mapped were the origin and termination of DNA replication. SV40 DNA replicates in productively infected cells via a circular intermediate, as illustrated in Fig 10 (13). Shown in the figure is an electron micrograph of the replicating SV40 DNA molecule where loops labeled L1 and L2 represent the replicated portion of the molecule, each loop containing one parental and one daughter DNA strand and the segment L3 represents the unreplicated portion of the molecule.

To determine whether SV40 DNA replicates from a unique site and whether replication occurs in

TABLE I
Order of *Hin* Fragments:
Analysis of Partial Digestion Products and *Hpa* Fragments

Initial Fragment (% of SV40 DNA)	*Hin* Digest Products	Overlapping Fragment Order										
12	G,J		J	G								
12	F,K	F										K
13	E,K										E	K
22	B,G			G	B							
40	B,F,G,J,K,	F	J	G	B							K
43	B,F,G,H,I,J	F	J	G	B	I	H					
51	A,C,D,E							A	C	D	E	
20 (*Hpa*-C)	B,I				B	I						
37 (*Hpa*-B)	A,H,C						H	A	C			
40 (*Hpa*-A)	D,E,F,G,J,K	F	J	G						D	E	K
	Order:	F	J	G	B	I	H	A	C	D	E	K

one direction or bidirectionally from that site a pulse-label experiment was carried out which was analogous to Dintzis' experiment on the biosynthesis of hemoglobin (14). The rationale of this experiment is illustrated in Fig 11. If one assumes that DNA replication does begin at a unique site and proceeds bidirectionally around the circular molecule, then, as illustrated in Fig 11, pulse labeling of replicating molecules with ^{3}H-thymidine would result in newly completed DNA molecules shown at the bottom of the figure which contain varying amounts of tritiated thymidine. Note that all molecules will contain tritium label at the site where DNA replication ends and few or no molecules will contain label at the site where DNA replication begins. Intermediate regions of the molecule will contain different amounts of label depending on their proximity to the origin. If we now cleave such a collection of newly completed, pulse-labeled molecules with a specific restriction endonuclease and determine the distribu-

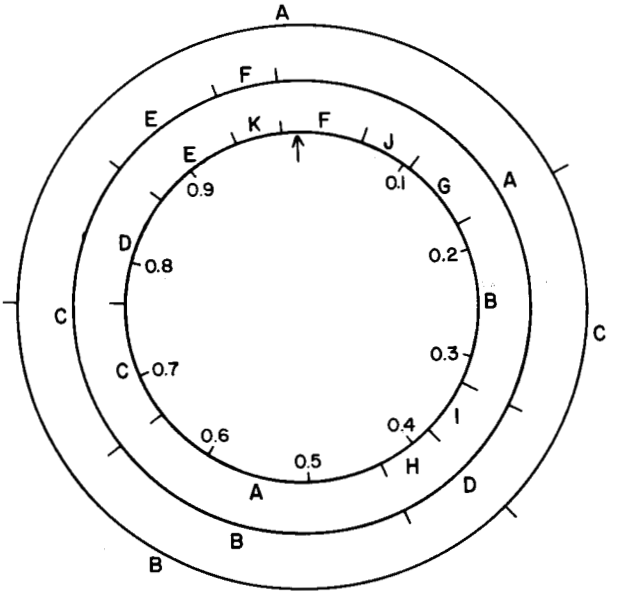

Fig. 8. *Hin* and *Hpa* cleavage maps of SV40 DNA. The arrow indicates the single *Eco*R$_1$ cleavage site. Outer circle, *Hpa* I fragments; middle circle, *Hin*d III fragments; inner circle, *Hin*dII + dIII fragments. Reproduced, with permission, from Lai and Nathans, Virology 66:72, 1975.

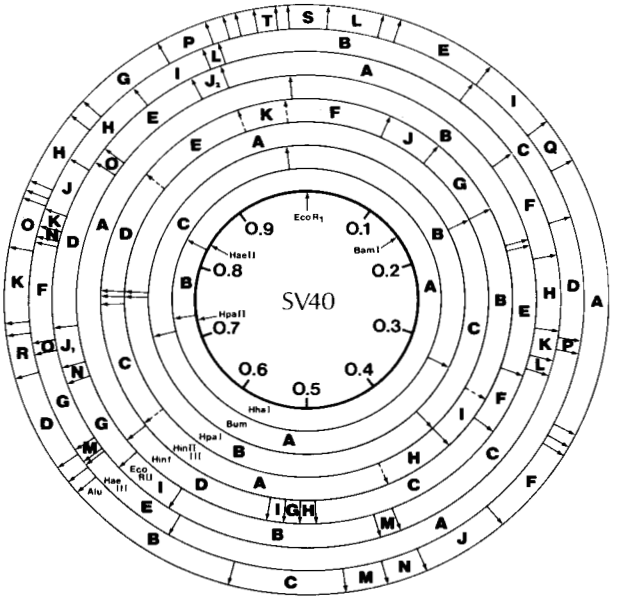

Fig. 9. Detailed cleavage map of the SV40 genome. Each circle represents cleavage sites (arrows) for the enzyme noted. (Courtesy of Cold Spring Harbor Laboratory.)

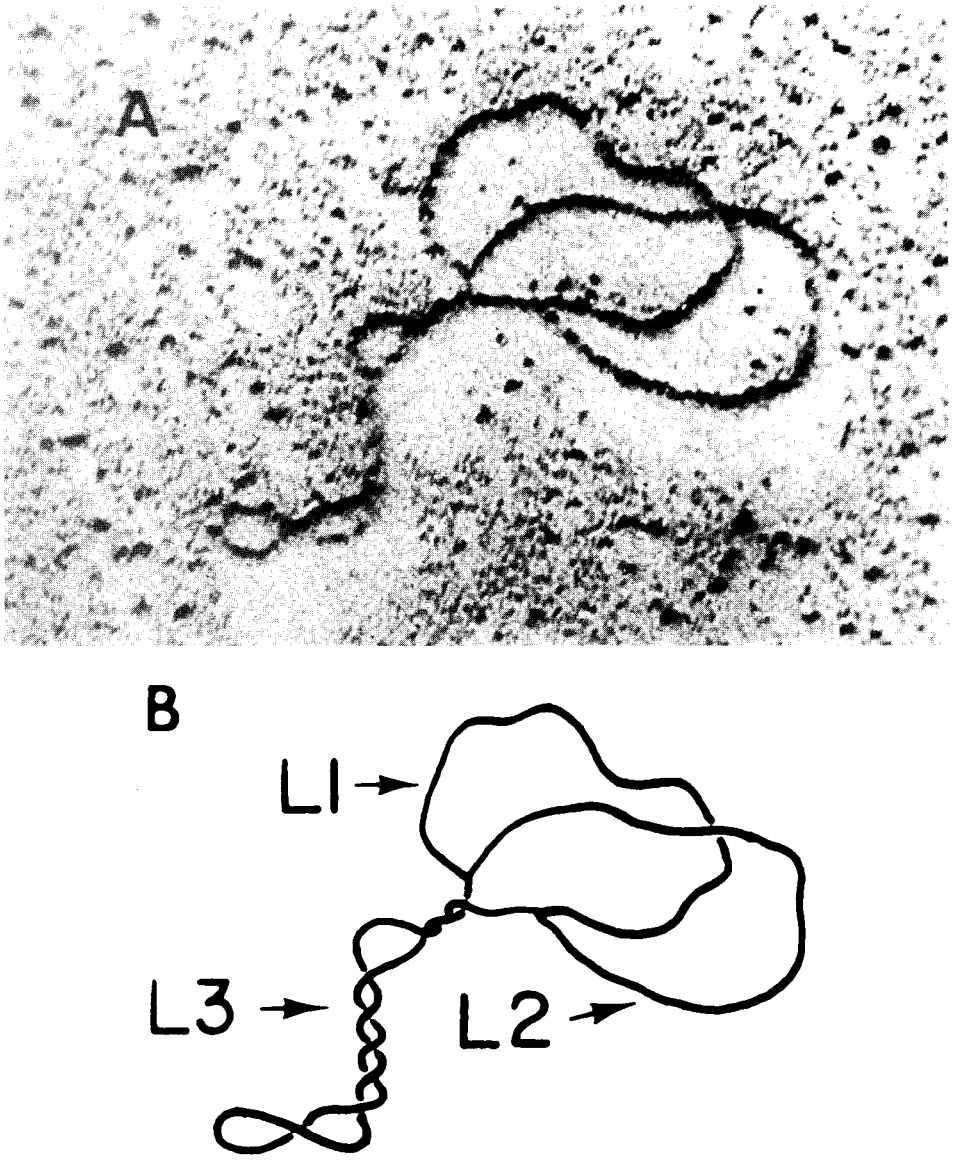

Fig. 10. SV40 DNA replicating intermediate. (A) Electron-micrograph of SV40 replicating intermediate, and (B) interpretive drawing. Replicative intermediates contain 2 forks, 3 branches, and no visible ends. Two of the branches (L1 and L2) are of equal length and represent the replicated segments of the molecule. The third branch (L3) contains superhelical twists and represents the unreplicated portion of the molecule. Figure courtesty of Dr. Thomas J. Kelly, Jr.

tion of pulse label in the fragments of DNA from different parts of the molecule, we should be able to deduce which part of the molecule was replicated first, which part was replicated last, and the order of replication of the segments between origin and terminus. This expectation is borne out by the experimental data obtained by K. J. Danna (42) shown in Fig 12, in which the amount of pulse label present in each *Hin* fragment is plotted against the position of that fragment within the genome for a pulse-labeling period of 5 min, 10 min, and 15 min. The data show that there are two gradients of pulse label within the DNA molecule starting at a site within *Hin*-C near the *Hin*-AC junction, which has the least amount of pulse label, and ending within fragment *Hin*-G on the opposite side of the DNA molecule, which has the most pulse label. We deduce from these results that there is a specific origin of SV40 DNA replication near the

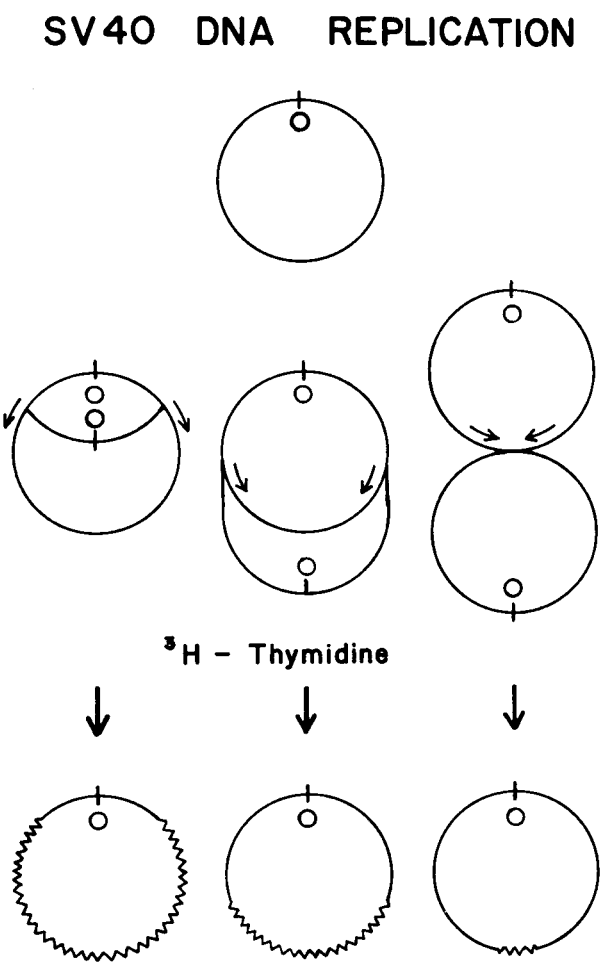

Fig. 11. Diagram of appearance of radioactive thymidine into newly replicated SV40 DNA molecules during a brief period of incorporation (less than a complete DNA replication cycle). Wavy lines indicate radioactive DNA segment. O represents the origin of replication. See text for discussion.

Hin-AC junction and that replication proceeds from that site at approximately equal rates and ends on the circular molecule opposite the origin. We presume then that SV40 DNA has a nucleotide sequence signal which determines where replication begins; this signal maps at about 0.67 map units on the SV40 cleavage map. A similar conclusion has been reached by Fareed et al. (15) on the basis of entirely different experiments.

The second template function of the SV40 genome which was mapped were segments corresponding to early and late messenger RNAs found in productivity infected or transformed cells. In experiments carried out by Khoury et al. (16, 17) and by Sambrook et al. (18), RNA isolated from SV40 infected or transformed cells was hybridized to individual strands of specific SV40 DNA fragments labeled with ^{32}P in order to determine which fragments contained nucleotide sequences corresponding to the SV40 RNA, as diagrammed in Fig 13. Those DNA strands which contained such sequences would form a DNA-RNA hybrid and thus could be separated by hydroxyapatite chromatography from those strands which did not contain sequences complementary to the RNA. In this way early RNA and late RNA from productively infected cells, and SV40 RNA found in a series of virus transformed cells lines were mapped on the SV40 cleavage map.

The results of a series of experiments of this type are shown in Fig 14. Also shown in the figure is

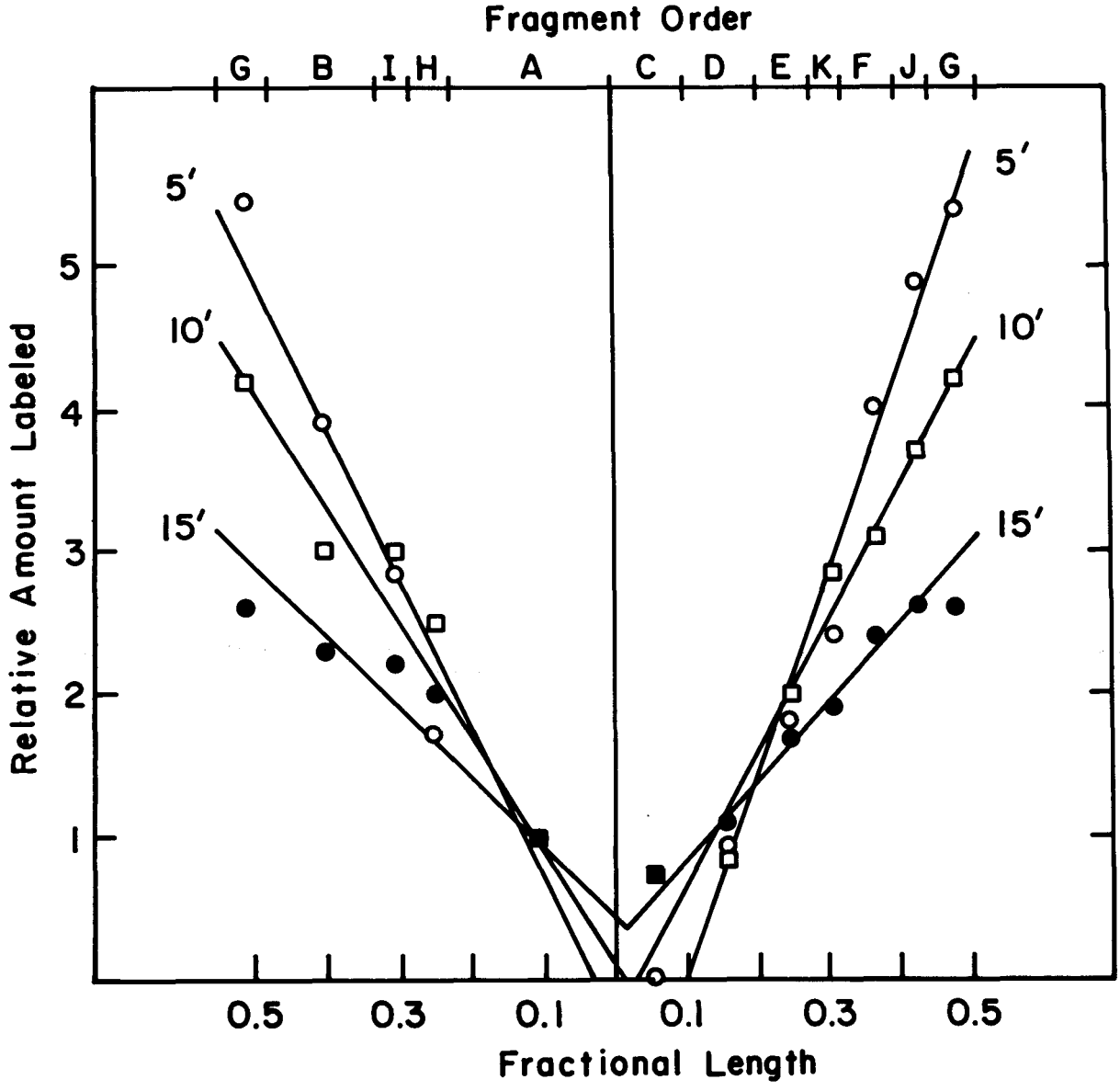

Fig. 12. Analysis of pulse-labeled, newly completed molecules of SV40 DNA. SV40-infected cells were pulse-labeled with ^{3}H-thymidine for 5, 10, or 15 min. Form I SV40 DNA was then isolated and mixed with uniformly labeled ^{32}P-SV40 DNA. The mixture was digested with endonuclease R·*Hin* and the $^{3}H/^{32}P$ ratio of each fragment measured. The normalized $^{3}H/^{32}P$ ratio is plotted on the ordinate vs the distance of the midpoint of each fragment from the A-C junction (0.655 map units). At the top of the figure is the order of fragments in the cleavage map, opened at fragment G, which is shown at each end. Reprinted, with permission, from Danna and Nathans (42).

the direction of early and late messenger RNA synthesis on the viral genome as determined in separate experiments, and those parts of the viral chromosome that correspond to two classes of late SV40 messenger RNAs found in infected cells, namely, 16S and 19S RNA, as recently reported by Khoury et al. (19) and by May et al. (20). The most striking finding was that the genome is divided into approximately two equal parts: one part extending from about 0.17 to 0.65 map units corresponds to the early region of the genome, and the other half corresponds to the late region of the genome. One of the dividing points between early and late regions is the origin of viral DNA replication. RNA from transformed cells corresponds rather closely to the early messenger RNA found in productively infected cells, except that in

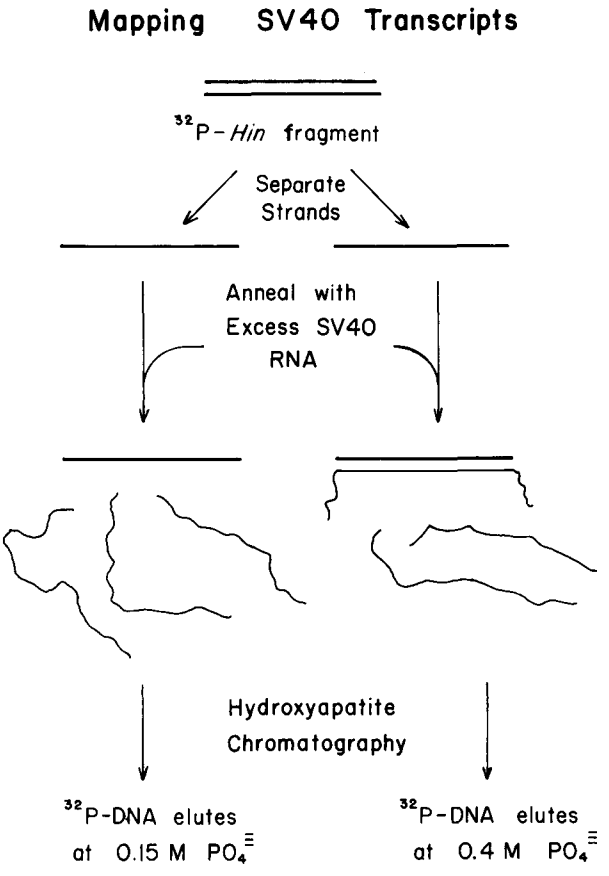

Fig. 13. Scheme for determining the segment and strand of SV40 DNA which is the template for SV40 RNA found in infected or transformed cells. Reprinted, with permission, from The Harvey Lectures, Series 69, p. 126, 1974–1975.

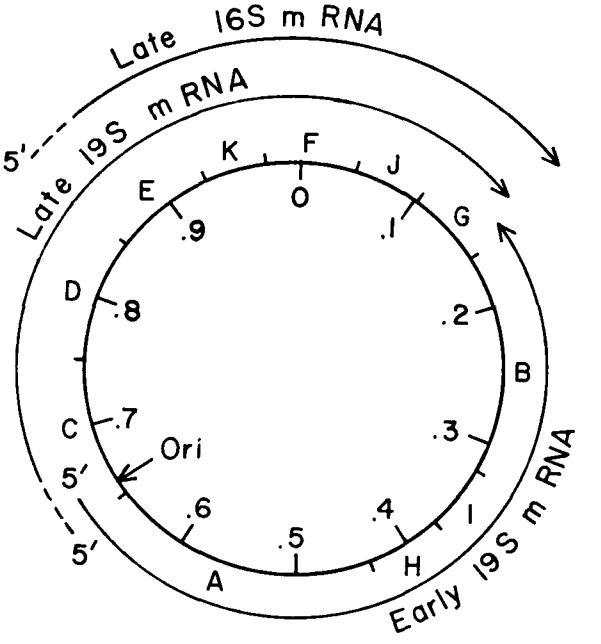

Fig. 14. Transcriptional map of SV40, in relation to the *Hin*d cleavage map. For details see the text. Reprinted, with permission, from Kelly and Nathans (1).

different transformed lines RNA sequences corresponding to *Hin* fragments C or C plus D are also present (21). These results confirm earlier findings that transformed cells contain predominantly early viral messenger RNA and in addition suggest that transcription of the viral messenger RNA may start in cellular DNA which is adjacent to the integrated viral genome in the cellular chromosome.

MAPPING STRUCTURAL GENES

To map structural genes one needs mutants with defined physiological defects. We have used two types of SV40 mutants in our mapping studies, temperature-sensitive mutants (referred to as ts mutants) and deletion mutants. Temperature-sensitive mutants of SV40 have been isolated and characterized by Tegtmeyer and Ozer (22), Kimura and Dulbecco (23), Chou and Martin (24), and by Dubbs et al. (25). Some of the findings from these laboratories are briefly summarized in Table II which presents the salient properties of different complementation classes of ts mutants. As seen in the table, there is only one class of mutant which is defective in viral DNA synthesis at high temperature, namely the tsA class. (tsD mutants which appear to be defective in viral DNA replication are actually defective in an uncoating property at high temperature; infection by viral DNA at high temperature is comparable to infection by wild type virus (26.) Mutants in the B, C, and BC classes are defective in a virion structural protein as shown by the fact that some mutants of this class produce temperature-sensitive virus particles. The two most important properties of tsA mutants are first, that at the nonpermissive temperature the initiation of viral DNA replication does not occur (27); and second, cells that are transformed by tsA mutants at permissive temperature revert to a nontransformed phenotype when shifted to the nonpermissive temperature (28). This last property together with a lack of transformation by tsA mutants at high temperatures, is illustrated in the diagram of Fig 15, and has been interpreted to indicate that functioning of the A gene product is required to initiate and maintain tumorigenic properties of transformed cells. Therefore, a specific viral gene producing a single gene product, called the A protein, appears to be required

TABLE II
Phenotypes of SV40 ts Mutants

Complementation Class	DNA Synthesis		ts	DNA	Antigens[a]				Transformation
	Virus	Host	Virions	Infection	T	U	C	V	
tsA	−	±	No	ts	+	−	−	−	−[b]
tsB	+	+	Yes	ts	+	+	+	+	+
tsC	+	+	Yes	ts					+
tsBC	+	+	Yes	ts	+	−	+	−	+
tsD	−	−	No	wt	−	−	−	−	−

[a] Antigens: T, T-antigen found early after infection; U, U-antigen also found early after infection—perhaps part of T-antigen molecule; C, capsid antigen, detected by antiserum against dissociated virus particles; V, virion antigen, detected by antiserum against intact virions.

[b] Initiation and maintenance of transformation are ts.

both for initiating viral DNA replication and for maintaining the transformed property of the cell.

In order to map temperature-sensitive mutants of SV40, C.-J. Lai adapted a method used in mapping bacteriophage mutants called marker rescue by fragments of DNA (29). In this method, which is illustrated in Fig 16, a single-stranded circle of mutant DNA is reacted with fragments from wild type viral DNA to produce a partial duplex molecule consisting of the fragment and circular DNA. Molecules of this type turn out to be infectious at 40°,

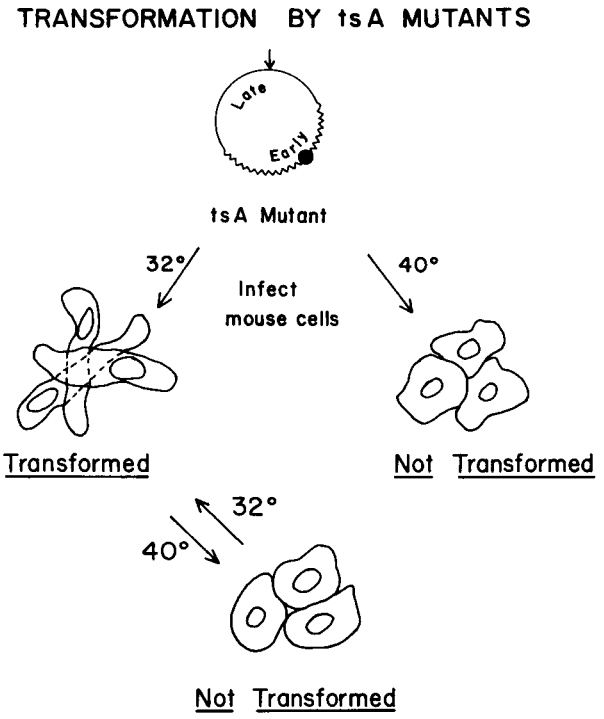

Fig. 15. Diagram of transformation by tsA mutants.

producing SV40 plaques in a monolayer of cells. Presumably within the infected cell the partial heteroduplex is converted to a complete heteroduplex molecule as illustrated in Fig 16, which then undergoes replication or correction of a mismatched base pair to give rise to wild type DNA. This chain of events will ensue to produce infectious virus if and only if the wild type fragment corresponds to the segment of the SV40 genome which contains the original mutational site, thereby allowing localization of the mutant site to that segment of the genome. In this way a number of temperature-sensitive mutants of SV40 have been mapped; the results are presented in Fig 17.

Our conclusions from this series of experiments can be summarized as follows. The mutants mapped cover about half of the SV40 genome; but there is notable absence of mutants mapping between 0.5 and 0.85 map units, and only one mutant which maps between 0.17 and 0.3 map units. Mutants which are defective in viral DNA synthesis and transformation (tsA mutants) all map in the early region of the genome. Mutants in complementation classes defective in a late function all map in the late region. B, C, and BC mutants are clustered in about one-half of the late region nearest the junction between the termination of early and late transcripts. D mutants all map between 0.85 and 0.94 map units, i.e., a small segment of the late region distinct from the B/C segment. On the basis of the mapping data and prior complementation tests we conclude that the B, C, and BC mutations are in one cistron, complementation occurring between B and C mutants at the protein level. D mutants are in a distinct cistron, as are the A mutants. The evidence thus suggests the presence of three SV40 genes: A, B/C, and D.

The second type of mutant used to map SV40 genes are deletion mutants. Such mutants have been obtained from two sources, one from viruses pas-

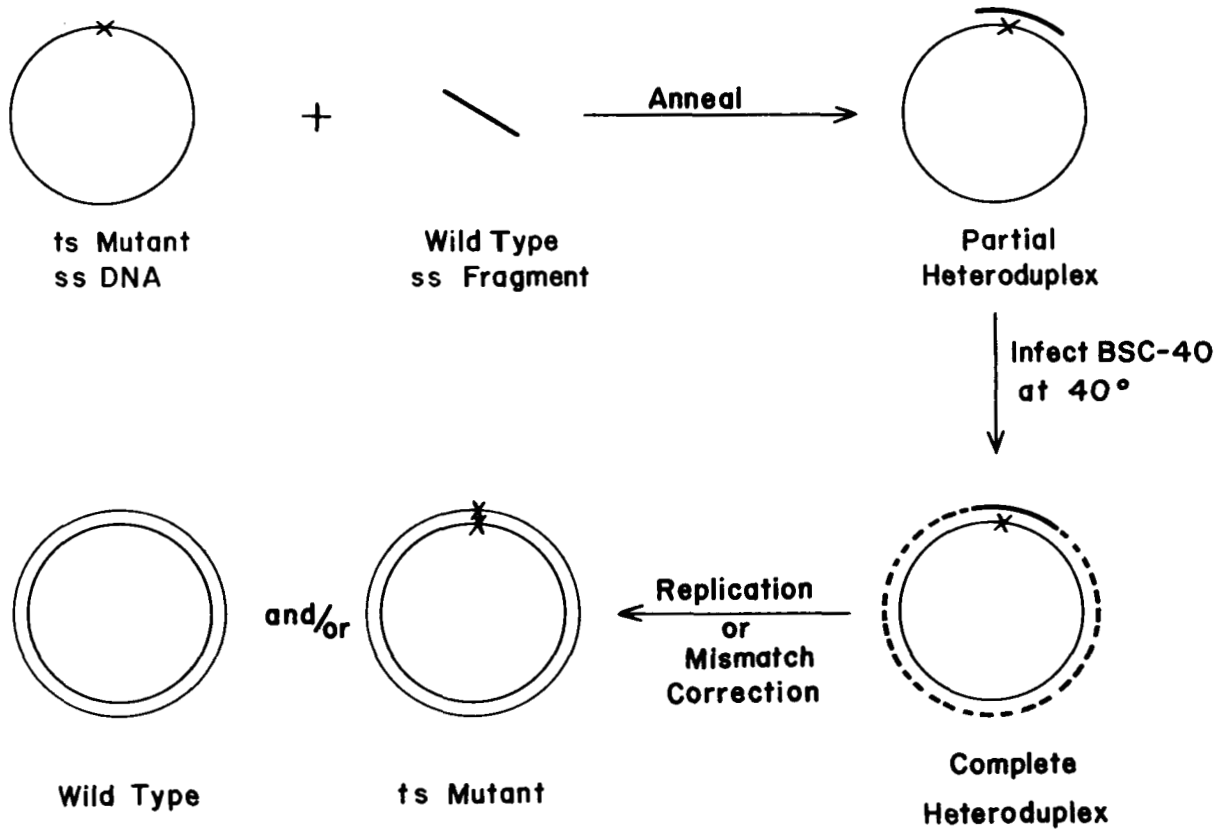

Fig. 16. Scheme for mapping ts mutants by marker rescue. X indicates the mutational site. Reprinted, with permission, from The Harvey Lectures, Series 69, p. 133, 1974–1975.

saged at high multiplicity of infection, under which conditions there is accumulation of defective virus particles containing deletions of viral DNA (30); and two, by enzymatic construction of mutants lacking specific DNA segments (31). Construction of deletion mutants is particularly fruitful since an essentially unlimited number of mutants can be isolated following enzymatic cleavage of DNA as illustrated in Fig 18. As shown in the figure following linearization of the circular DNA molecule with or without enzymatic excision of a segment of DNA by a given restriction enzyme or nonspecific endonuclease, linear DNA can be used to infect cells and resulting deletion mutants cloned and identified. Since the deletion mutants are defective, they must be propagated and cloned in the first place in the presence of a suitable helper virus by a procedure called complementation plaque formation developed by Brockman and Nathans (32); and by Mertz and Berg (33), see Fig 19. It is interesting to note that any linear DNA molecule taken up by the cells cyclizes intracellularly by recombination near the ends of

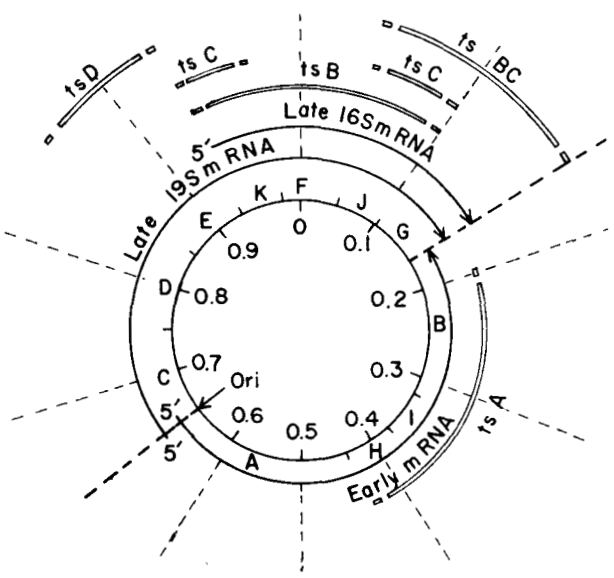

Fig. 17. A map of ts mutants of SV40.

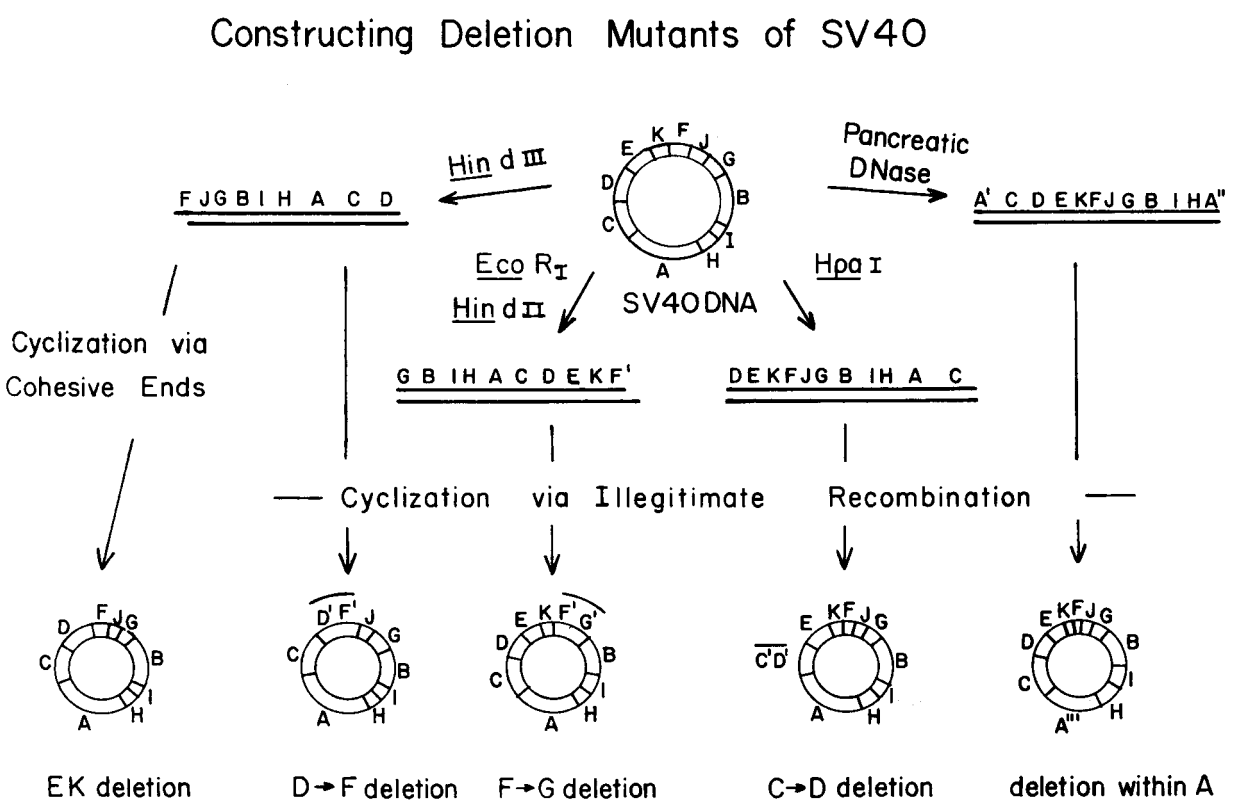

Fig. 18. Generation of SV40 deletion mutants by the use of specific or nonspecific endonuclease to generate linear molecules followed by complementation cloning. Reprinted, with permission, from The Harvey Lectures, Series 69, p. 146, 1974–1975. See text for discussion.

such a linear molecule to produce a covalently closed circular DNA (31). Evidently, the monkey cells used in these experiments contain an active enzymatic system which can carry out this type of "illegitimate" recombination, i.e., recombination not dependent on extensive base sequence homology at the recombination sites.

Brockman, Lai, and Adler have isolated a large number of deletion mutants of this type and have mapped them by restriction endonuclease cleavage followed by electrophoresis of fragments and by electron microscopic heteroduplex mapping (31). For example, as shown in Fig 20, deletion mutant No. 2 has a missing segment of DNA detected by electrophoresis of the *Hin* fragment produced from this DNA relative to that produced from wild type DNA. By electron microscopic heteroduplex mapping as shown in Fig 20, this mutant DNA was found to have a deletion corresponding to the missing *Hin* fragment, and by appropriate measurement of heteroduplex molecules the deletion can be precisely localized in the cleavage map. Examples of other deletion mutants of this type are shown in Fig 21. As seen in the figure, some mutants contain deletions in the early region of the SV40 genome while others contain deletions in the late region.

Having a series of mutants of this type, we could use them to correlate physiological defects with the map positions of the deletions (34). To illustrate, deletion mutants were tested by Scott and Brockman for their ability to transform mouse cells as illustrated in Fig 22. As seen in the figure, a deletion mutant lacking part of the early region of the SV40 genome but retaining the entire late region in functional form is unable to transform cells. In contrast, a deletion mutant lacking nearly the entire late region but containing the early region of SV40 intact is able to transform (and, in fact, with efficiency equal to that of wild type virus). Moreover, from such transformed cells Walter Scott has been able to rescue by appropriate cell fusion techniques the original mutants virus, thus confirming that the transformation was caused by the mutant genome.

Experiments similar to those just described have also been carried out with fragments of SV40 DNA by van der Eb and his colleagues (35). In their experiments a fragment of SV40 DNA between map unit 0.15 and map unit 0.74, which includes the entire

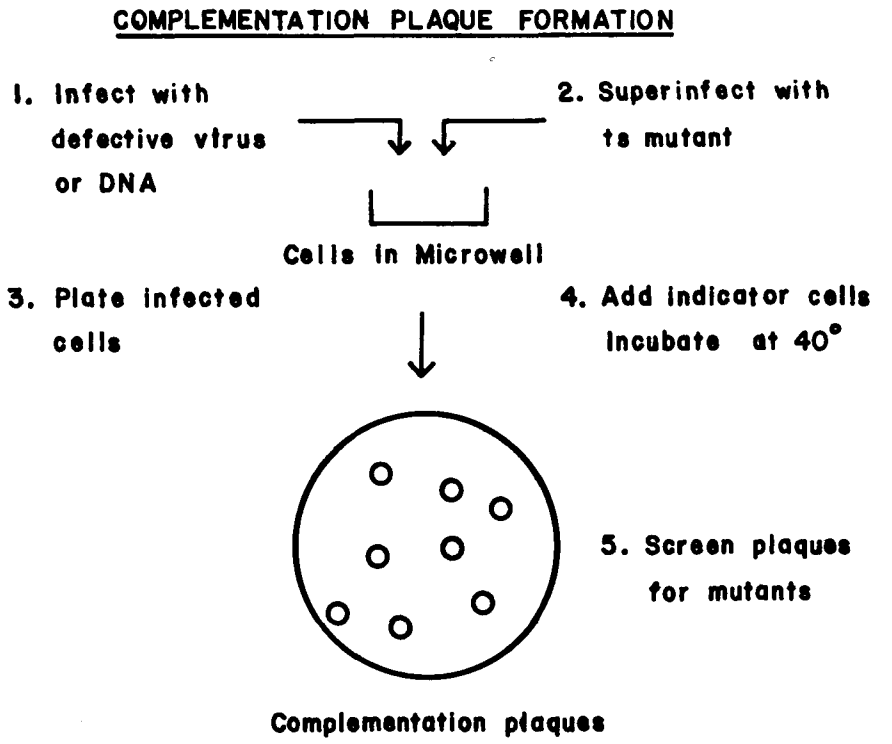

Fig. 19. Method of cloning SV40 variants by complementation with ts mutants. See text for details. Reprinted, with permission, from Kelly and Nathans (1).

early region of the genome can transform rat cells at an efficiency approximately equal to that of the complete DNA molecule, thus establishing that the early region plus contiguous segments at each end is sufficient to transform. All of these results are, of course, consistent with the prior finding that tsA mutants fail to transform cells at nonpermissive temperature, and taken together they indicate that the early region of SV40 is required for transformation and is also sufficient at least when DNA segments on either side of the early region are also present.

PRIMARY SV40 GENE PRODUCTS

Deletion mutants of SV40 have also been used to identify primary gene products, that is proteins which are coded for by the SV40 DNA. The rationale of these experiments is that a deletion within a gene may give rise to a short polypeptide chain coded for by that gene (Fig 23). With this in mind, deletion mutants of SV40 were used to infect cells, and proteins made during infection were labeled with ^{35}S methionine and subsequently separated by gel electrophoresis and identified by radioautography (41). We were looking for new polypeptides not present in cells infected by wild type virus nor in uninfected cells. Since in virus infected cells, the major structural protein of the virus (VP1) is easily identified, Ching-Juh Lai began with a deletion mutant suspected of having its deletion in the VP1 gene. This mutant, shown in Fig 21 as *dl*-1010, has a deletion between map units 0.99 and 0.11, extending from *Hin* fragment F to *Hin* fragment G. When cells were infected with this mutant virus and labeled proteins separated by electrophoresis and radioautography, a new polypetide was detected which by tryptic peptide analysis was shown to be related to VP1. We therefore concluded that the VP1 gene maps in this segment of the late region of the SV40 genome and since this segment contains tsB, C, and BC mutants (Fig 17), the VP1 gene corresponds to the B/C gene. Independent evidence that this region of the genome codes for VP1 has been obtained by translation of the messenger RNA derived the segment of the genome in the VP1 polypeptide *in vitro* by Prives et al. (36) and Lodish et al. (37), and also by recent comparison of amino acid sequences of the N-terminus of VP1 with the nucleotide sequence within *Hin* fragment K by Fiers and his colleagues.

A similar type of experiment has been carried out by Randell et al. (38) with one of our mutants containing a deletion in the early region of the

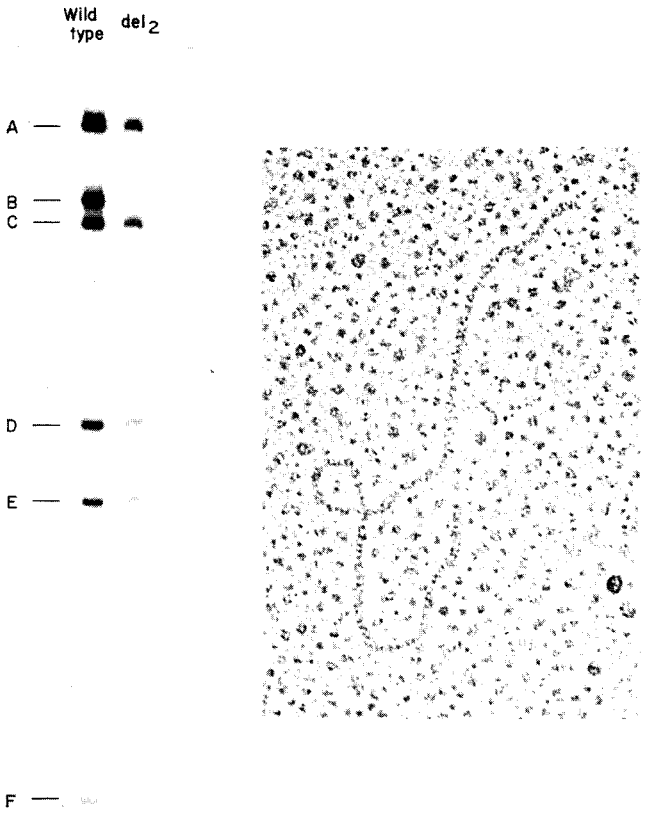

Fig. 20. *Left*: Endo R·*Hin* digest of DNA of a constructed deletion mutant (*dl*-1002) missing one *Hin* fragment (fragment B).

Right: A heteroduplex DNA molecule seen in the electron microscope. One strand of the duplex is wild type SV40 DNA-$_{LR_I}$ (i.e., linear DNA made by endo R·*Eco*R$_I$ cleavage) and the other strand is *dl*-1002 DNA-$_{LR_I}$. The single strand loop is the part of wild type DNA opposite the missing segment of DNA in the *dl*-1002 strand.

genome between map units 0.32 and 0.43 encompassing *Hin* fragments H and I, corresponding to the region where most tsA mutants map (see *dl*-1001, Fig 21). Identification of the early gene product in this case depended on isolation of the SV40 specific T or tumor antigen from infected cells (39). Tegtmeyer showed that radioactively labeled proteins from SV40 infected cells when precipitated by anti-T serum yielded a polypeptide in the precipitate of about 100,000 daltons as determined by electrophoretic mobility in SDS polyacrylamide gel (39). When similar experiments were carried out with the early deletion mutant just described, a new and shorter polypeptide precipitable by anti-T serum was detected, from which it was inferred that the deletion caused the formation of a shorter T polypeptide than that produced in wild type SV40-infected cells. Therefore, the T-antigen polypeptide is the primary gene product of the early region. Since T-antigen found in wild type SV40-infected cells is about 100,000 daltons, requiring about 2500 nucleotides to code for it, it is likely that the entire early region is required to code for this single polypeptide.

From the experiments just described and prior experiments on complementation between mutants of SV40 as well as the results of translation of specific viral messenger RNAs, it appears likely that the SV40 genome contains three distinct genes: the early A gene which codes for a polypeptide of about 100,000 daltons containing T-antigen determinants; a late B/C gene which codes for the major virion structural protein VP1; and a late D gene which probably

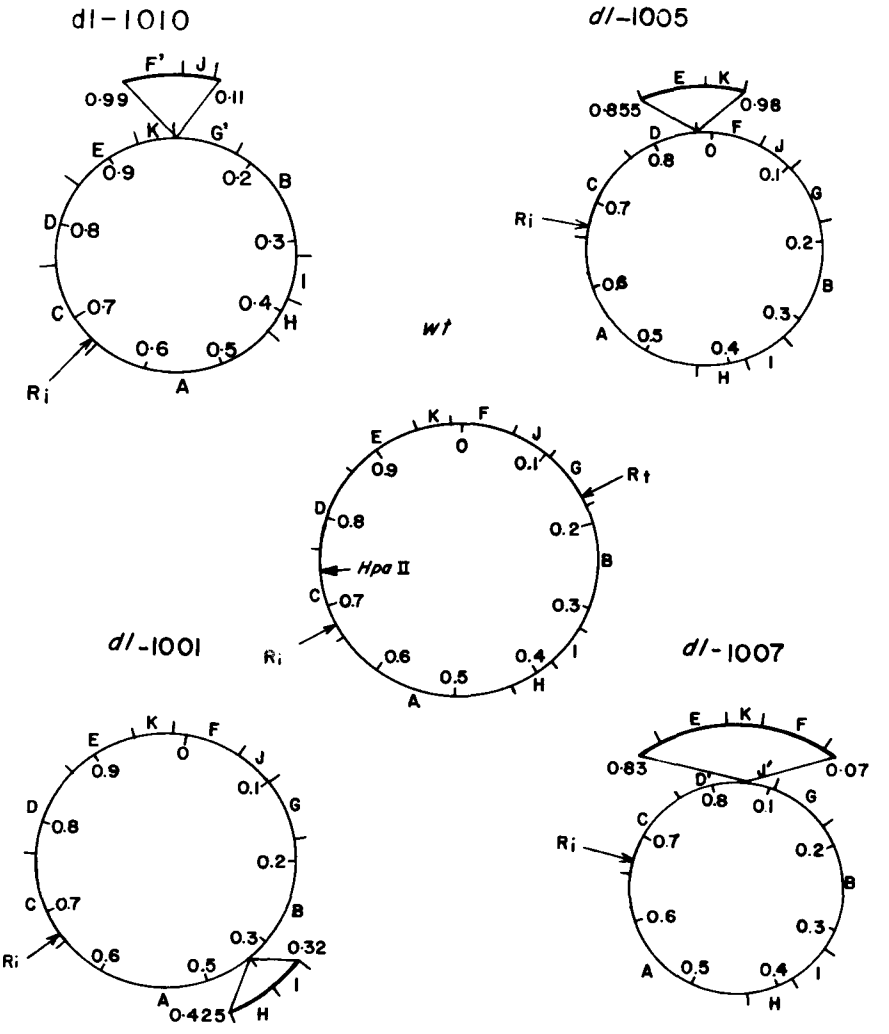

Fig. 21. Cleavage maps of several deletion mutants constructed as described in the text. Deleted regions are indicated by the wedged segments.

codes for a minor structural protein. The possibility that there are other small polypeptides coded for by the viral genome has not been entirely excluded.

SUMMARY MAP OF THE SV40 GENOME

All of the data which have been presented on the genetic organization of SV40 can be summarized in a physiological map shown in Fig 24. To summarize the salient features of the map, there is a unique origin of viral DNA replication, which proceeds bidirectionally from the origin and terminates approximately half way around the circle; the early and late regions of the genome are continuous stretches of DNA each occupying approximately half of the genome; the 5′ ends of stable early and late large messenger RNA's map near the origin of DNA replication and this may be the structural correlate of the linkage between replication and late gene transcription; there are three identifiable structural genes of the virus, A, B/C, and D, the last two being late genes that code for structural proteins, the first an early gene coding for a single polypeptide chain, which contains determinants for T-antigen. Also shown are the sites where determinants of U-antigen, T-antigen, and tumor specific transplantation antigen of SV40 have been more precisely localized by analysis of adeno-SV40 hybrid viruses (40).

THE A GENE AND TRANSFORMATION

The segment of the genome required for cell

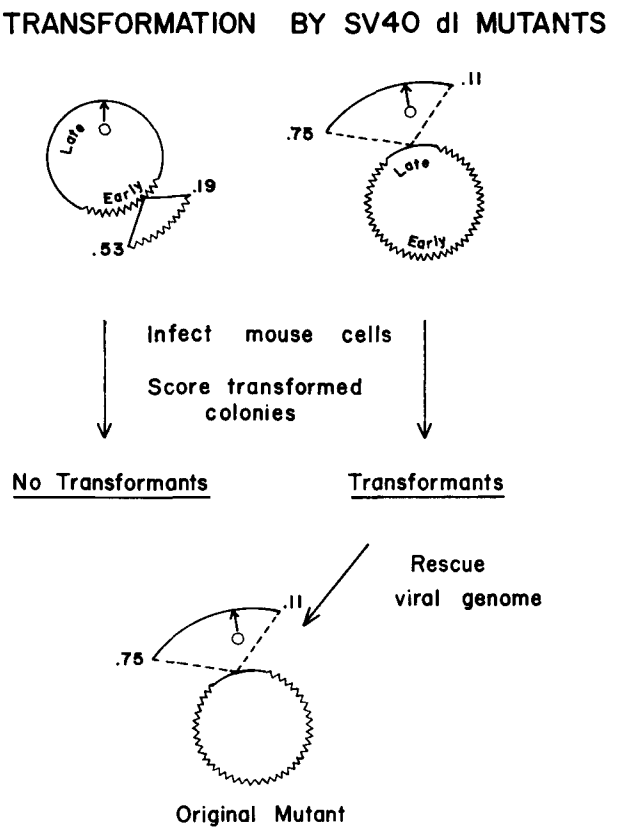

Fig. 22. Diagram of transformation activity of mutants with large deletions in the early or late segment of the genome. As seen at the bottom, the original deletion mutant could be rescued from transformed cells.

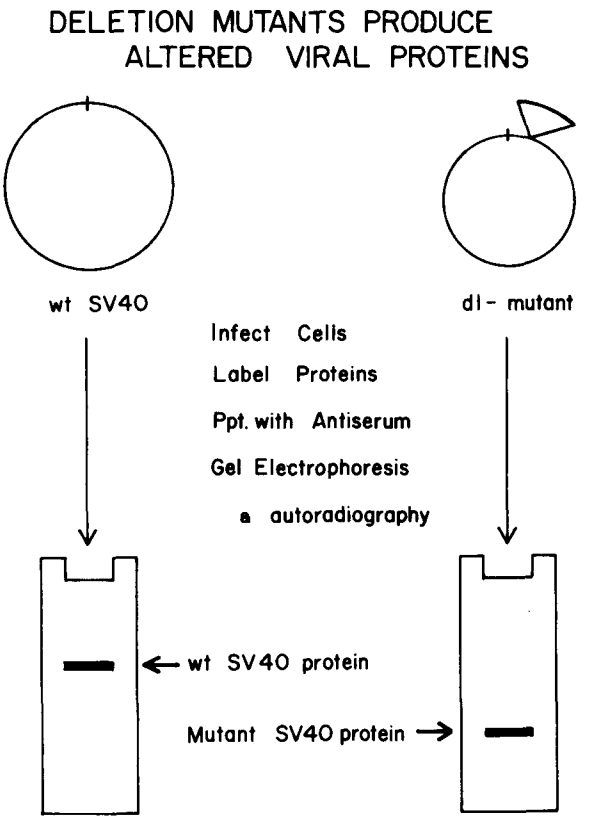

Fig. 23. Diagram of procedure for detecting short SV40-specified proteins in cells infected with a deletion mutant.

transformation by SV40 A gene is indicated in the physiological map of Fig 24, based on the experiments described earlier. How this segment of DNA causes transformation remains the most important question concerning viral tumorigenesis. Since the primary gene product of this region, the A-protein, is required for cell transformation, the function of the A-protein is of great interest. Genetic evidence indicates that the A-protein is involved in initiation of viral DNA replication (27), and biochemical experiments suggest that the A-protein can bind to viral DNA, perhaps specifically at the site of origin of DNA replication (43). Based on these observations it has been suggested that the A-protein similarly recognizes nucleotide sequences in cell DNA corresponding to the SV40 DNA origin, and thereby initiates DNA replication at unusual sites in cellular DNA, as illustrated in Fig 25. If transcription of cellular genes is somehow coupled to cellular DNA replication, as is transcription of the late genes of SV40, such aberrant initiation of cell DNA replication may cause the

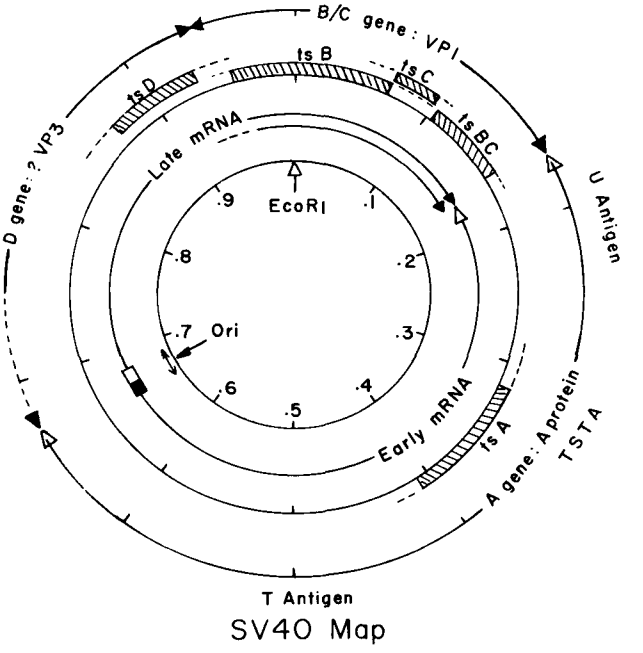

Fig. 24. Summary map of the SV40 genome. Reproduced, with permission, from Kelly and Nathans (1).

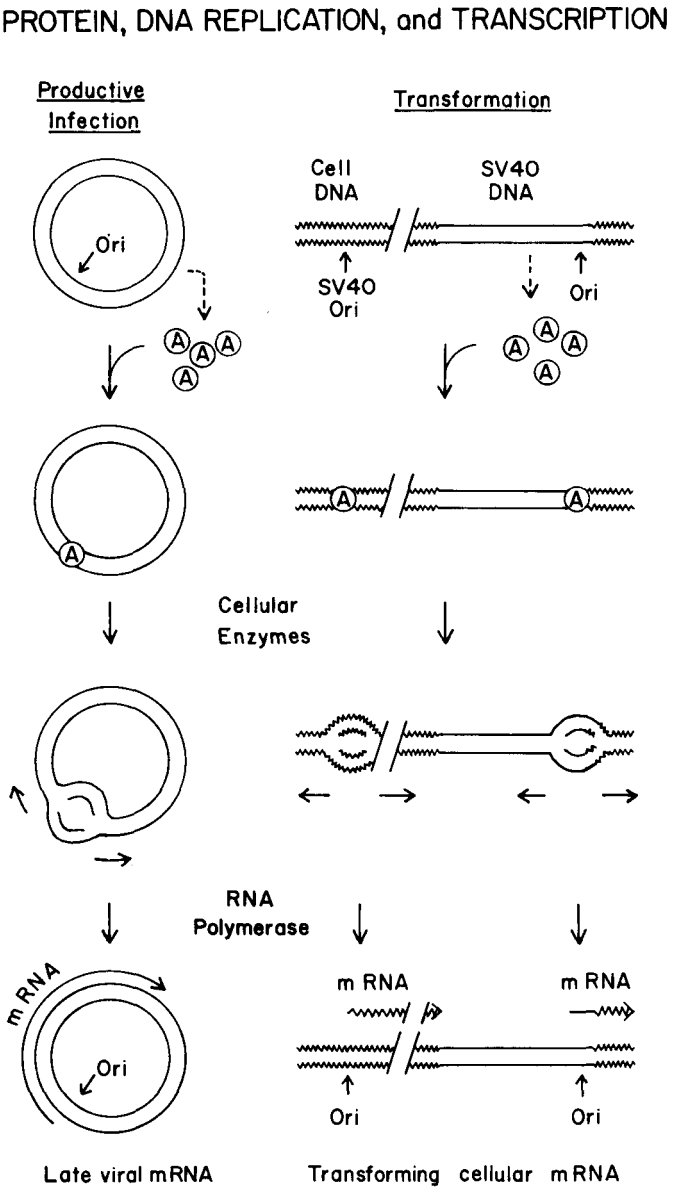

Fig. 25. Possible involvement of SV40 early gene product (A protein) in viral DNA replication and late gene transcription in productively infected cells (left) and in cellular DNA replication and transcription in transformed cells.

expression of genes previously quiescent. These gene products could be the proximate cause of cell transformation. To test this hypothesis will require a great deal of work. Hopefully an understanding of the molecular genetics of SV40 and related viruses will provide experimental tools to help analyze the molecular events in the transformed cell that are responsible for its tumorigenic properties.

ACKNOWLEDGMENT

It is a pleasure to acknowledge the primary contributions of my coworkers: Stuart Adler, Kathleen Danna, George Sack, William Brockman, Theresa Lee, Mary Gutai, Ching-Juh Lai, Walter Scott, Nicholas Muzyczka, and David Shortle. I am also grateful for the constant advice and criticism of Drs. Hamilton O. Smith and Thomas J. Kelly, Jr.

REFERENCES

1. Kelly Jr, TJ and Nathans D: The genome of simian virus 40. Adv Virus Res, 1976, in press.
2. Sweet BH and Hilleman, MR: The vacuolating virus, SV_{40}. Proc Soc Exp Biol Med 105:420, 1960
3. Anderer FA et al.: Structure of simian virus 40, II. Symmetry and components of the virus particle. Virology 32:511, 1967
4. Gerry HW, Kelly Jr. TJ, and Berns KI: Arrangement of nucleotide sequences in adeno-associated virus DNA. J Mol Biol 79:207, 1973
5. Kelly Jr. TJ and Smith HO: A restriction enzyme from *Hemophilus influenzae,* II. Base sequence of the recognition site. J Mol Biol 51:393, 1970
6. Smith HO and Wilcox KW: A restriction enzyme from *Hemophilus influenzae,* I. Purification and general properties. J Mol Biol 51:379, 1970
7. Adler SP and Nathans D: Cleavage of simian virus 40 DNA by bacterial restricting enzymes. Fed Proc (Abs) 29:725, 1970
8. Nathans D and Smith HO: Restriction endonucleases in the analysis and restructuring of DNA molecules. Ann Rev Biochem 44:273, 1975
9. Danna KJ and Nathans D: Specific cleavage of simian virus 40 DNA by restriction endonuclease of *Hemophilus influenzae.* Proc Nat Acad Sci USA 68:2913, 1971
10. Danna KJ, Sack Jr. GH, and Nathans D: Studies of SV40 DNA, VII. A cleavage map of the SV40 genome. J Mol Biol 78:363, 1973
11. Morrow JF and Berg P: Cleavage of SV40 DNA at a unique site by a bacterial restriction enzyme. Proc Nat Acad Sci USA 69:3365, 1972
12. Mulder C and Delius H: Specificity of the break produced by endonuclease R_I in simian virus 40 DNA, as revealed by partial denaturation mapping. Proc Nat Acad Sci USA 69:3215, 1972
13. Sebring ED et al.: Structure of replicating simian virus 40 deoxyribonucleic acid molecule. J Virol 8:478, 1971
14. Dintzis HM: Assembly of the peptide chains of hemoglobin. Proc Nat Acad Sci USA 47:247, 1961
15. Fareed GC, Garon CF, and Salzman NP: Origin and direction SV40 deoxyribonucleic acid replication. J Virol 10:484, 1972
16. Khoury G et al.: A map of SV40 transcription sites expressed in productively infected cells. J Mol Biol 78:377, 1973
17. Khoury G et al.: Posttranscriptional selection of simian virus 40-specific RNA. J Virol 15:433, 1975.
18. Sambrook J et al.: Genome localization of simian virus 40 RNA species. J Virol 17:832, 1976
20. May E, Kopecka H, and May P: Mapping the transcription site of the SV40-speciific late 16S mRNA. Nucleic Acids Res 2:1995, 1975
21. Khoury G et al.: A transcription map of the SV40 genome in transformed cell lines. Virology 63:263, 1975
22. Tegtmeyer P and Ozer HJ: Temperature sensitive mutants of simian virus 40: Infection of permissive cells. J Virol 8:516, 1971
23. Kimura G and Dulbecco R: Isolation and characterization of temperature-sensitive mutants of simian virus 40. Virology 49:394, 1972
24. Chou JY and Martin RG: Complementation analysis of simian virus 40 mutants. J Virol 13:1101, 1974
25. Dubbs DR, Rachmeler M, and Kit S: Recombination between temperature-sensitive mutants of simian virus 40. Virology 57:161, 1974
26. Robb JA and Martin RG: Genetic analysis of SV40, III. Characterization of temperature sensitive mutant blocked at an early stage of productive infection in monkey cells, J Virol 9:956, 1972
27. Tegtmeyer P: SV40 DNA synthesis. The viral replicon. J Virol 10:591, 1972
28. Brugge JS and Butel JS: Role of simian virus 40 gene A function in maintenance of transformation. J Virol 15:619, 1975
29. Lai C-J and Nathans D: Mapping temperature-sensitive mutants of simian virus 40: Rescue of mutants by fragments of viral DNA. Virology 60:466, 1974
30. Yoshiike K: Studies on DNA from low-density particles of SV40, I. Heterogeneous defective virions produced by successive undiluted passages. Virology 34:391, 1968
31. Lai C-J and Nathans D: Deletion mutants of simian virus 40 generated by enzymatic excision of DNA segments from the viral genome. J Mol Biol 89:179, 1974
32. Brockman WW and Nathans D: The isolation of simian virus 40 variants with specifically altered genomes. Proc Nat Acad Sci USA 71:942, 1974
33. Mertz JE and Berg P: Defective simian virus 40 genomes: Isolation and growth of individual clones. Virology 62:112, 1974
34. Scott WA, Brockman WW, and Nathans D: Biological activities of deletion mutants of simian virus 40. Virology, in press
35. Abrahams PJ et al.: Transformation of primary rat kidney cells by fragments of simian virus 40 DNA. J Virol 16:818, 1975
36. Prives CL et al.: Cell-free translation of messenger RNA of simian virus 40: Synthesis of the major capsid protein. Proc Nat Acad Sci USA 71:302, 1974
37. Lodish HF, Weinberg RA, and Ozer HL: Translation of mRNA from SV40-infected cells into SV40 capsid protein by cell-free extracts. J Virol 13:590, 1974
38. Rundell K et al.: Identification of simian virus 40 protein A. J. Virol, 1976, in press
39. Tegtmeyer P: Altered patterns of protein synthesis in infection by SV40 mutants. Cold Spring Harbor Symp Quant Biol 39:9, 1974
40. Lewis Jr. AM et al.: Studies of hamster cells transformed by adenovirus 2 and the nondefective Ad2-SV40 hybrids. Cold Spring Harbor Symp Quant Biol 39:651, 1974
41. Lai C-J and Nathans D: The B/C gene of simian virus 40. Virology, 1976, in press
42. Danna KJ and Nathans D: Studies of SV40 DNA, IV. Bidirectional replication of simian virus 40 DNA. Proc Nat Acad Sci USA 69:3097, 1972
43. Reed SI et al.: T antigen binds to simian virus 40 DNA

at the origin of replication. Proc Nat Acad Sci USA 72:1605, 1975
44. Old RW, Murray K, and Roizes G: Recognition sequence of restriction endonuclease III from *Hemophilus influenzae*. J Mol Biol 92:331, 1975
45. Garfin DE and Goodman HM: Nucleotide sequences at the cleavage sites of two restriction endonucleases from *Hemophilus parainfluenzae*. Biochem Biophys Res Com 59:108, 1974
46. Hedgpeth J, Goodman HM, and Boyer HW: DNA nucleotide sequence restricted by the RI endonuclease. Proc Nat Acad Sci US 69:3448, 1972
47. Boyer HW et al.: DNA substrate site for the Eco_{RII} restriction endonuclease and modification methylase. Nature New Biol 244:40, 1973
48. Bigger CH, Murray K, and Murray NE: Recognition sequence of a restriction enzyme. Nature New Biol 244:7, 1973